

بسم الله الرحمن الرحيم

مبانی پردازش گفتار و کاربردهای بازشناسی

تصنیف

منصور شیخان

دانشگاه آزاد اسلامی - واحد تهران جنوب



۱۳۹۳

سرشناسه: شیخان، منصور، ۱۳۴۴-

عنوان و نام پدیدآور: مبانی پردازش گفتار و کاربردهای بازشناسی / تصنیف منصور شیخان.

مشخصات نشر: تهران: دانشگاه آزاد اسلامی، واحد تهران جنوب، ۱۳۹۳.

مشخصات ظاهری: ۴۵۱ ص.؛ جدول، نمودار.

شابک: ۹۷۸-۹۶۴-۱۰-۳۲۰۴-۵

قیمت: ۳۲۰۰۰۰ ریال

وضعیت فهرست‌نویسی: فیا

یادداشت: ص. ع. به انگلیسی: Principles of Speech Processing and Recognition Applications

یادداشت: کتابخانه: ص. ۳۲۴-۳۵۴.

موضوع: گفتار - ترکیب، گفتارپرداز، زبان‌شناسی کامپیوتری، پردازش زبان طبیعی، فارسی - داده‌پرداز.

شناسه افزوده: دانشگاه آزاد اسلامی، واحد تهران جنوب.

رده‌بندی کنگره: ۱۳۹۳ش۹/۷۸۸۲/TK

رده‌بندی دیویی: ۰۰۶/۵۴

شماره کتاب‌شناسی ملی: ۳۷۰۶۷۱۱



مبانی پردازش گفتار و کاربردهای بازشناسی

تصنیف منصور شیخان

ناشر: دانشگاه آزاد اسلامی - واحد تهران جنوب

رئیس انتشارات: فرشته حاجی‌زاده

ویراستار: دکتر نعمت طالبی

طراح جلد: سید محمد شاهی لنگرودی

حروف‌چینی و صفحه‌آرایی: امیرحسین فاضلی

چاپ اول: ۱۳۹۳

تیراژ: ۲۰۰۰ نسخه

قیمت: ۲۲۰۰۰ تومان

لیتوگرافی، چاپ و صحافی: سازمان چاپ دانشگاه آزاد اسلامی

شابک: ۹۷۸-۹۶۴-۱۰-۳۲۰۴-۵

حق چاپ برای ناشر محفوظ است.

بسم الله الرحمن الرحيم

«ن والقلم و ما یسطرون»

سخن ناشر

بی تردید تعالی علمی و فرهنگی کشور، بازتاب میزان تفکری است که فرهیختگان در مقام عمل به منصه ظهور می‌رسانند. در این بین، مهم‌تر و حساس‌تر از همه، رسالت عالمان و علم‌جویان است. این رسالت، همان گونه که در جهت پرورش استعدادهای فردی، با قرائت و مطالعه انجام می‌گیرد، برای پیشبرد اهداف اجتماعی و تعالی فرهنگی، با کتابت و مباحثه نیز محقق می‌شود.

از این رو، یکی از مسئولیت‌های مهم مراکز فرهنگی در کنار تعلیم و تربیت، نشر آثار علمی است. معاونت پژوهشی دانشگاه آزاد اسلامی - واحد تهران جنوب، برای ارتقای سطح علمی دانشجویان و تنویر افکار دانش‌پژوهان، به چاپ کتاب **مبانی پردازش گفتار و کاربردهای بازشناسی** اقدام کرده است؛ با این امید که با استمرار فعالیت‌های انتشاراتی، بتواند گامی در راه رشد و بالندگی فرهنگی بردارد.

این کتاب در دوره آموزشی مقاطع کارشناسی‌ارشد و دکترای مهندسی برق (در گرایش‌های مخابرات و الکترونیک) و مقاطع کارشناسی‌ارشد و دکترای مهندسی پزشکی (در گرایش بیوالکتریک) به عنوان یکی از مراجع درس **پردازش گفتار** و نیز مقاطع کارشناسی‌ارشد و دکترای هوش مصنوعی به عنوان یکی از مراجع درس **پردازش و شناسایی گفتار** کاربرد دارد.

پیش از هر چیز، شایسته است از عنایات ریاست محترم دانشگاه آزاد اسلامی - واحد تهران جنوب، استادان ارجمند و نیز از دست‌اندرکاران ویرایش، حروف‌چینی و چاپ کتاب، تقدیر و تشکر به عمل آید. از صاحب‌نظران گرامی خواهشمند است با ارائه پیشنهادهای خود، در جهت اصلاح نواقص احتمالی این کتاب و تهیه مطالب مناسب متون درسی دانشگاهی، انتشارات این واحد را یاری کنند. امید است این اثر علمی مورد استفاده استادان و دانش‌پژوهان قرار گیرد.

و من الله التوفیق و علیه التکلان

دانشگاه آزاد اسلامی - واحد تهران جنوب

فهرست مطالب

۱۹	پیشگفتار
۲۳	۱. مبانی و کاربردهای پردازش گفتار
۲۳	۱-۱. مقدمه
۲۵	۲-۱. زمینه‌های کاربردی سامانه‌های پردازش گفتار
۲۶	۱-۲-۱. سامانه‌های کدکردن گفتار
۲۶	۲-۲-۱. سامانه‌های تبدیل متن به گفتار
۳۰	۳-۲-۱. سامانه‌های بازشناسی گفتار
۳۳	۴-۲-۱. سامانه‌های بازشناسی گوینده
۳۵	۵-۲-۱. سامانه‌های بازشناسی حالت از روی گفتار
۳۶	۶-۲-۱. سامانه‌های بازشناسی نوع زبان
۳۸	۷-۲-۱. سامانه‌های بازشناسی جنس گوینده
۳۸	۸-۲-۱. سامانه‌های بهسازی گفتار
۳۹	۹-۲-۱. سامانه‌های تبدیل گفتار
۴۰	۳-۱. برخی از فعالیت‌های انجام‌شده در زمینه‌های کاربردی سامانه‌های پردازش گفتار در کشور

۴۵	۴-۱. جمع‌بندی فصل
۴۵	۵-۱. پرسش‌ها
۴۹	۲. مبانی تولید و درک سیگنال گفتار
۴۹	۱-۲. مقدمه
۴۹	۲-۲. تولید گفتار و بازنمایی آوایی آن
۵۳	۳-۲. انتشار صوت و مدل لوله بدون افت یکنواخت
۵۵	۴-۲. انتشار موج در مدل لوله‌های اتصال‌یافته بدون افت
۵۸	۵-۲. مدل تولید گفتار
۶۰	۶-۲. سازوکار شنوایی در انسان
۶۲	۷-۲. نظریه باندهای بحرانی در شنوایی
۶۴	۸-۲. پدیده پوشش در شنوایی
۶۵	۹-۲. جمع‌بندی فصل
۶۵	۱۰-۲. پرسش‌ها
۶۷	۱۱-۲. مسائل
۶۹	۳. تحلیل‌های کوتاه‌مدت سیگنال گفتار
۶۹	۱-۳. مقدمه
۷۱	۲-۳. چند نمونه از پارامترهای پایه در تحلیل زمانی کوتاه‌مدت
۷۲	۱-۲-۳. انرژی کوتاه‌مدت
۷۲	۲-۲-۳. تابع میانگین اندازه کوتاه‌مدت
۷۲	۳-۲-۳. نرخ عبور از صفر کوتاه‌مدت
۷۴	۴-۲-۳. AMDF کوتاه‌مدت
۷۴	۳-۳. تابع خودهمبستگی کوتاه‌مدت

۷۷	۴-۳. تبدیل فوریۀ کوتاهمدت (STFT)
۷۷	۴-۳.۱. تعریف و تفاسیر مختلف برای STFT
۷۸	۴-۳.۲. نمونه برداری از STFT در حوزه های زمان و فرکانس
۷۹	۴-۳.۳. طیف نگاشت گفتار
۸۲	۴-۳.۴. رابطه بین STFT و STACF
۸۲	۴-۳.۵. سنتز فوریۀ کوتاهمدت
۸۴	۴-۳.۵. تعیین کوتاهمدت دورۀ تناوب گام
۸۸	۴-۳.۶. تمایز گفتار از سکوت
۹۰	۴-۳.۷. جمع بندی فصل
۹۱	۴-۳.۸. پرسش ها
۹۳	۴-۳.۹. مسائل
۹۷	۴. تحلیل همومرفیک سیگنال گفتار
۹۷	۴-۱. مقدمه
۹۹	۴-۲. چگونگی محاسبۀ کپستروم
۱۰۰	۴-۲-۱. به کارگیری تبدیل فوریۀ گسسته در محاسبه
۱۰۰	۴-۲-۲. به کارگیری تبدیل Z در محاسبه
۱۰۲	۴-۲-۳. محاسبات بازگشت پذیر کپستروم مختلط
۱۰۳	۴-۳. پالایش همومرفیک کوتاهمدت سیگنال گفتار
۱۰۵	۴-۴. کاربرد تحلیل همومرفیک در آشکارسازی گام
۱۰۵	۴-۵. کاربرد در بازشناسی الگوهای گفتاری
۱۰۵	۴-۵-۱. مبانی و جبران سازی اثرات اعوجاج خطی
۱۰۷	۴-۵-۲. معیارهای فاصلۀ مبتنی بر کپستروم لیقترشده

۱۰۷	۴-۶. ضرایب کپستروم حوزه فرکانسی مل (MFCC)
۱۰۹	۴-۷. نسخه‌های اصلاحی برای پارامترهای MFCC
۱۱۰	۴-۸. افزایش مقاومت در برابر نویز در سامانه‌های بازشناسی مبتنی بر MFCC
۱۱۱	۴-۹. کاربرد پارامترهای مبتنی بر تحلیل همومرفیک در برخی از زمینه‌های پردازش گفتار
۱۱۱	۴-۹-۱. کاربرد ضرایب کپستروم در بازشناسی گفتار
۱۱۳	۴-۹-۲. کاربرد ضرایب کپستروم در بازشناسی گوینده
۱۱۳	۴-۹-۳. کاربرد ضرایب کپستروم در بازشناسی حالت از روی گفتار
۱۱۴	۴-۹-۴. کاربرد ضرایب کپستروم در برخی از سایر حوزه‌های پردازش گفتار
۱۱۵	۴-۱۰. جمع‌بندی فصل
۱۱۶	۴-۱۱. پرسش‌ها
۱۱۷	۴-۱۲. مسائل
۱۲۰	۵. تحلیل پیش‌بینی خطی
۱۲۰	۵-۱. مقدمه
۱۲۰	۵-۲. ارتباط پیش‌بینی خطی و مدل گفتار
۱۲۴	۵-۳. روش‌های تعیین ضرایب پیش‌بینی
۱۲۴	۵-۳-۱. روش کوواریانس و روال بازگشت‌پذیر آن
۱۲۸	۵-۳-۲. روش خودهمبستگی و روال بازگشت‌پذیر آن
۱۳۴	۵-۴. طیف حاصل از تحلیل پیش‌بینی خطی
۱۳۷	۵-۵. بازنمایی‌های معادل برای ضرایب LPC
۱۳۷	۵-۵-۱. ریشه‌های تابع سامانه فیلتر خطای پیش‌بینی
۱۳۸	۵-۵-۲. ضرایب LSP
۱۳۹	۵-۵-۳. کپستروم پاسخ ضربه لوله صوتی

۱۴۰	 ضرایب PARCOR ۴-۵-۵
۱۴۰	 ضرایب لگاریتم نسبت سطح مقطع‌ها (LAR) ۵-۵-۵
۱۴۱	 مدل پیش‌بینی خطی ادراکی (PLP) ۶-۵
۱۴۳	 کاربرد LPC و نسخه‌های اصلاحی در کدکردن گفتار ۷-۵
۱۴۵	 کاربرد پارامترهای LPC در برخی از زمینه‌های پردازش گفتار ۸-۵
۱۴۵	 کاربرد ضرایب LPC و نسخه‌های اصلاحی در بازشناسی گفتار ۱-۸-۵
۱۴۶	 کاربرد ضرایب LPC و نسخه‌های اصلاحی در بازشناسی گوینده و حالت ۲-۸-۵
۱۴۷	 کاربرد ضرایب LPC و نسخه‌های اصلاحی در برخی از سایر حوزه‌های پردازش گفتار ۳-۸-۵
۱۴۸	 جمع‌بندی فصل ۹-۵
۱۴۹	 پرسش‌ها ۱۰-۵
۱۵۱	 مسائل ۱۱-۵
۱۵۶	 سامانه‌های کدکردن گفتار ۶-۵
۱۵۶	 مقدمه ۱-۶
۱۵۶	 نمونه‌برداری و چندی‌سازی گفتار ۲-۶
۱۵۹	 تحلیل نوبز چندی‌سازی یکنواخت ۱-۲-۶
۱۶۰	 چندی‌سازی بر اساس قاعدهٔ μ ۲-۲-۶
۱۶۰	 چندی‌سازی غیریکنواخت و وقفی ۳-۲-۶
۱۶۲	 کدکردن گفتار رقمی‌شده ۳-۶
۱۶۴	 کدکننده‌های حلقه- بسته ۴-۶
۱۶۴	 کدکردن مبتنی بر پیش‌بینی ۱-۴-۶
۱۶۷	 مدوله‌سازی دلتا ۲-۴-۶
۱۶۸	 سامانه‌های ADPCM ۳-۴-۶

- ۱۶۹ ADPCM ۱-۳-۴-۶. کدکردن پارامترهای
- ۱۷۱ ADPCM ۲-۳-۴-۶. کیفیت برحسب نرخ بیت برای کدکننده‌های
- ۱۷۲ AbS ۴-۴-۶. کدکردن به روش
- ۱۷۲ AbS ۱-۴-۴-۶. سامانه کدکردن پایه به روش
- ۱۷۳ ۲-۴-۴-۶. وزن دهی ادراکی سیگنال تفاضلی
- ۱۷۴ ۳-۴-۴-۶. تولید سیگنال تحریک
- ۱۷۵ ۴-۴-۴-۶. کدکردن به روش پیش‌بینی خطی چند-پالسی
- ۱۷۷ CELP ۵-۴-۴-۶. کدکردن به روش
- ۱۸۰ AbS ۶-۴-۴-۶. پیش‌بینی کننده‌های تأخیر زیاد در کدکننده‌های
- ۱۸۱ ۵-۶. کدکننده‌های حلقه- باز
- ۱۸۱ ۱-۵-۶. مدل با تحریک دو-حالتی
- ۱۸۱ V/UV ۱-۱-۵-۶. تعیین گام، بهره و تصمیم
- ۱۸۲ ۲-۱-۵-۶. تخمین پارامترهای سامانه لوله صوتی
- ۱۸۴ ۲-۵-۶. کدکردن به روش پیش‌بینی خطی با تحریک توسط سیگنال مانده (RELp)
- ۱۸۶ ۳-۵-۶. سامانه‌های با تحریک مخلوط
- ۱۸۶ MELP ۱-۳-۵-۶. کدکننده
- ۱۹۰ HVXC ۲-۳-۵-۶. کدکننده از نوع
- ۱۹۳ MBE ۳-۳-۵-۶. کدکننده
- ۱۹۸ ۶-۶. کدکننده‌های ادراکی گفتار
- ۱۹۸ ۱-۶-۶. پردازش شنیداری گفتار
- ۱۹۹ ۱-۱-۶-۶. ساختار کلی کدکننده ادراکی گفتار
- ۱۹۹ ۲-۱-۶-۶. پوشش فرکانسی و زمانی

۲۰۰	۲-۶-۶. ملاحظات کد کردن ادراکی
۲۰۰	۱-۲-۶-۶. محدودیت‌های قابلیت تفکیک زمان / فرکانس
۲۰۱	۲-۲-۶-۶. مدل MBE برای کد کردن ادراکی
۲۰۱	۷-۶. جمع‌بندی فصل
۲۰۲	۸-۶. پرسش‌ها
۲۰۴	۹-۶. مسائل
۲۰۹	۷. سامانه‌های بازشناسی خودکار گفتار
۲۰۹	۱-۷. مقدمه
۲۱۰	۲-۷. مدل پایه برای ASR
۲۱۱	۳-۷. تحلیل ویژگی با هدف بازنمایی مقاوم از سیگنال گفتار
۲۱۲	۴-۷. پیاده‌سازی یک سامانه ASR
۲۱۳	۱-۴-۷. مجموعه ویژگی برای بازشناسی
۲۱۴	۲-۴-۷. موضوع بازشناسی
۲۱۴	۳-۴-۷. آموزش در بازشناسی
۲۱۴	۴-۴-۷. آزمون و ارزیابی عملکرد
۲۱۵	۵-۷. فرایندهای تصمیم‌گیری در ASR
۲۱۵	۱-۵-۷. فرمول‌سازی ریاضی مسأله
۲۱۶	۲-۵-۷. مدل مخفی مارکف
۲۱۹	۳-۵-۷. مدل‌سازی آکوستیکی
۲۲۲	۴-۵-۷. مدل زبانی
۲۲۳	۵-۵-۷. مسأله جستجو
۲۲۴	۶-۵-۷. کاربرد شبکه‌های عصبی مصنوعی در ASR

۲۲۵	 ۷-۵-۱. دسته‌بندی ایستای قطعات گفتاری
۲۳۱	 ۷-۵-۲. دسته‌بندی پویای قطعات گفتاری
۲۳۴	 ۷-۵-۷. سامانه‌های آمیختار HMM/ANN
۲۳۵	 ۷-۶. بازشناسی گفتار دارای نوا و حالت
۲۳۸	 ۷-۷. مقاوم‌سازی سامانه‌های ASR
۲۳۹	 ۷-۷-۱. روش‌های جبران‌سازی سیگنال
۲۳۹	 ۷-۷-۲. روش‌های فضای ویژگی
۲۴۰	 ۷-۷-۲. روش‌های فضای مدل
۲۴۱	 ۷-۷-۴. به‌کارگیری محاسبات نرم در موضوع مقاوم‌سازی بازشناسی
۲۴۲	 ۷-۸. برخی از پژوهش‌های انجام‌شده در حوزه ASR در کشور
۲۵۱	 ۷-۹. جمع‌بندی فصل
۲۵۱	 ۷-۱۰. پرسش‌ها
۲۵۶	۸. سامانه‌های بازشناسی گوینده و جنسیت و سن گویندگان
۲۵۶	 ۸-۱. مقدمه
۲۵۷	 ۸-۲. مبانی سامانه‌های بازشناسی گوینده
۲۵۸	 ۸-۲-۱. انتخاب ویژگی‌ها
۲۵۹	 ۸-۲-۲. مدل‌سازی گوینده
۲۵۹	 ۸-۳. استخراج ویژگی‌ها
۲۵۹	 ۸-۳-۱. ویژگی‌های طیفی کوتاه‌مدت
۲۶۰	 ۸-۳-۲. ویژگی‌های منبع صوتی
۲۶۰	 ۸-۳-۳. ویژگی‌های زمانی - طیفی
۲۶۳	 ۸-۳-۴. ویژگی‌های نوایی

۲۶۴	۸-۳-۵. ویژگی‌های سطح بالا
۲۶۴	۸-۴-۴. روش‌های متداول برای مدل‌سازی گوینده
۲۶۴	۸-۴-۱. چندی‌سازی برداری (VQ)
۲۶۵	۸-۴-۲. مدل مخلوط گوسی (GMM)
۲۶۹	۸-۴-۳. ماشین بردار پشتیبان (SVM)
۲۷۰	۸-۴-۴. سایر مدل‌ها
۲۷۱	۸-۴-۵. ترکیب مدل‌ها
۲۷۲	۸-۵-۵. سامانه‌های بازشناسی گوینده مقاوم در برابر نویز
۲۷۲	۸-۵-۱. هنجارسازی ویژگی‌ها
۲۷۳	۸-۵-۲. جبران‌سازی مدل گوینده
۲۷۳	۸-۵-۳. هنجارسازی امتیاز
۲۷۴	۸-۶-۶. روش‌های ابربردار
۲۷۴	۸-۶-۱. مبانی ابربردار
۲۷۶	۸-۶-۲. SVM با هسته GLDS
۲۷۷	۸-۶-۳. SVM با ابربردار گوسی
۲۷۷	۸-۶-۴. SVM با ابربردار MLLR
۲۷۸	۸-۶-۵. SVM با ابربردار سطح بالا
۲۷۸	۸-۶-۶. هنجارسازی ابربردارهای SVM
۲۷۹	۸-۶-۷. روش‌های تحلیل عامل
۲۸۱	۸-۷-۷. سامانه‌های بازشناسی جنسیت و سن گویندگان
۲۸۱	۸-۷-۱. مبانی و کاربردها
۲۸۲	۸-۷-۲. نمونه‌هایی از سامانه‌های بازشناسی جنسیت و سن گویندگان

- ۲۸۲ ۸-۷-۱. سامانه مبتنی بر بازشناسی های واج موازی
- ۲۸۳ ۸-۷-۲. سامانه مبتنی بر MLP-DBN
- ۲۸۴ ۸-۷-۳. سامانه مبتنی بر HMM
- ۲۸۵ ۸-۷-۴. سامانه مبتنی بر GMM
- ۲۸۵ ۸-۷-۵. سامانه مبتنی بر GMM-UBM
- ۲۸۶ ۸-۸. برخی پژوهش های انجام شده در حوزه بازشناسی گوینده و جنسیت / سن گوینده در کشور
- ۲۸۸ ۸-۹. جمع بندی فصل
- ۲۸۹ ۸-۱۰. پرسش ها
- ۲۹۳ ۹. سامانه های بازشناسی حالت احساسی گوینده
- ۲۹۳ ۹-۱. مقدمه
- ۲۹۵ ۹-۲. دادگان گفتار دارای حالت احساسی
- ۲۹۵ ۹-۲-۱. معیار طراحی دادگان
- ۲۹۶ ۹-۲-۲. برخی از دادگان موجود برای گفتار دارای حالت احساسی
- ۲۹۹ ۹-۲-۳. مشکلات دادگان موجود برای گفتار دارای حالت احساسی
- ۳۰۰ ۹-۳. ویژگی های گفتاری مناسب برای سامانه های ESR
- ۳۰۰ ۹-۳-۱. ویژگی های محلی و سراسری
- ۳۰۱ ۹-۳-۲. دسته های مختلف ویژگی های گفتاری
- ۳۰۱ ۹-۳-۲-۱. ویژگی های گفتاری پیوسته نوایی
- ۳۰۳ ۹-۳-۲-۲. ویژگی های کیفی صوت
- ۳۰۴ ۹-۳-۲-۳. ویژگی های گفتاری مبتنی بر طیف
- ۳۰۵ ۹-۳-۲-۴. ویژگی های غیرخطی مبتنی بر TEO
- ۳۰۷ ۹-۳-۳. پردازش گفتار در سامانه های ESR

۳۰۸	 روش‌های انتخاب ویژگی نمونه برای ESR
۳۰۹	 ۱-۴-۳-۹. معیارهای مبتنی بر همبستگی
۳۱۰	 FCBF روش ۲-۴-۳-۹
۳۱۲	 ANOVA روش ۳-۴-۳-۹
۳۱۷	 ۵-۳-۹. ترکیب ویژگی‌های آکوستیکی با سایر منابع اطلاعاتی
۳۱۷	 ۱-۵-۳-۹. ترکیب اطلاعات آکوستیکی و زبانی
۳۱۸	 ۲-۵-۳-۹. ترکیب اطلاعات آکوستیکی، زبانی و موضوع گفتار
۳۱۸	 ۳-۵-۳-۹. ترکیب اطلاعات آکوستیکی و دیداری
۳۱۹	 ۴-۹. روش‌های دسته‌بندی در سامانه‌های ESR
۳۱۹	 HMM ۱-۴-۹
۳۲۱	 GMM ۲-۴-۹
۳۲۱	 ۳-۴-۹. شبکه عصبی
۳۲۲	 SVM ۴-۴-۹
۳۲۴	 ۵-۴-۹. سامانه‌های دسته‌بندی چندتایی
۳۲۷	 ۵-۹. جمع‌بندی فصل
۳۲۸	 ۶-۹. پرسش‌ها
۳۳۳	 ۱۰. سامانه‌های بازشناسی خودکار نوع زبان
۳۳۳	 ۱-۱۰. مقدمه
۳۳۵	 ۲-۱۰. مبانی سامانه‌های LID
۳۳۶	 ۳-۱۰. ویژگی‌های گفتاری مورد استفاده در سامانه‌های LID
۳۳۷	 ۱-۳-۱۰. ویژگی‌های آکوستیکی
۳۳۷	 ۲-۳-۱۰. اطلاعات واج‌آرایی

- ۳۳۸ ۱۰-۳-۳. اطلاعات نوایی
- ۳۳۸ ۱۰-۳-۴. ریخت‌شناسی لغوی
- ۳۳۸ ۱۰-۳-۵. دستور زبان
- ۳۳۹ ۱۰-۴. روش‌های مطرح در بخش استخراج ویژگی‌های آکوستیکی در سامانه‌های LID
- ۳۴۰ ۱۰-۴-۱. ضرایب دلتا- کیستروم و دلتا- دلتا- کیستروم
- ۳۴۱ ۱۰-۴-۲. ضرایب دلتا- کیستروم جابه‌جاشده (SDC)
- ۳۴۲ ۱۰-۴-۳. ویژگی‌های مبتنی بر مدوله‌سازی فرکانس (FM)
- ۳۴۵ ۱۰-۵. روش‌های مطرح در بخش دسته‌بندی از سامانه‌های LID
- ۳۴۵ ۱۰-۵-۱. GMM
- ۳۴۶ ۱۰-۵-۲. GMM-UBM
- ۳۴۷ ۱۰-۵-۳. GMM-SVM
- ۳۴۹ ۱۰-۵-۴. GLDS-SVM
- ۳۵۰ ۱۰-۵-۵. تاریخچه سایر روش‌های دسته‌بندی مورد استفاده در سامانه‌های LID
- ۳۵۰ ۱۰-۶. روش‌های هنجارسازی
- ۳۵۱ ۱۰-۶-۱. تفریق میانگین کیستروم (CMS)
- ۳۵۱ ۱۰-۶-۲. پیش‌فشار ویژگی
- ۳۵۲ ۱۰-۶-۳. VTLN
- ۳۵۳ ۱۰-۶-۴. روش NAP
- ۳۵۴ ۱۰-۷. سامانه‌های LID مبتنی بر اطلاعات واج‌آرایی
- ۳۵۴ ۱۰-۷-۱. مدل‌سازی زبان پس از بازشناسی واج (PRLM)
- ۳۵۵ ۱۰-۷-۲. مدل‌سازی زبان پس از بازشناسی واج موازی
- ۳۵۵ ۱۰-۷-۳. ترکیب مدل‌سازی فضای بردار و PPR

۳۵۸	۸-۱۰. سامانه‌های LID مبتنی بر اطلاعات نوایی، ریخت‌شناسی و دستور زبان
۳۵۸	۱۰-۸-۱. به‌کارگیری اطلاعات نوایی
۳۵۸	۱۰-۸-۲. به‌کارگیری اطلاعات ریخت‌شناسی و دستور زبان
۳۵۹	۱۰-۹. گروه‌بندی زبان‌ها و دسته‌بندی تسلسلی
۳۵۹	۱۰-۹-۱. پیش‌دسته‌بندی مبتنی بر نواخت
۳۵۹	۱۰-۹-۲. بازشناسی تسلسلی زبان
۳۶۲	۱۰-۱۰. دادگان مورد استفاده در حوزه LID
۳۶۲	۱۰-۱۰-۱. دادگان گفتار تلفنی OGI
۳۶۳	۱۰-۱۰-۲. دادگان گفتار تلفنی CallFriend
۳۶۳	۱۰-۱۰-۳. دادگان NIST LRE
۳۶۴	۱۰-۱۱. ارزیابی‌های NIST
۳۶۵	۱۰-۱۲. بررسی عملکرد سامانه‌های نمونه در حوزه LID
۳۶۶	۱۰-۱۲-۱. سامانه LID نمونه پیشنهادی آزمایشگاه Lincoln از MIT
۳۶۶	۱۰-۱۲-۱-۱. بخش پردازش ویژگی‌های طیفی
۳۶۷	۱۰-۱۲-۱-۲. سامانه‌های بازشناسی شباهت طیفی
۳۶۷	۱۰-۱۲-۱-۳. سامانه‌های مبتنی بر نشان‌گذاری
۳۶۸	۱۰-۱۲-۲. سامانه‌های مبتنی بر اطلاعات آکوستیکی
۳۶۹	۱۰-۱۲-۳. سامانه‌های مبتنی بر اطلاعات واج‌آرایی
۳۷۰	۱۰-۱۲-۴. سامانه‌های مبتنی بر اطلاعات نوایی
۳۷۱	۱۰-۱۲-۵. عملکرد برخی از سامانه‌های LID با ساختار دسته‌بندی‌کننده‌های مبتنی بر اطلاعات واج‌آرایی
۳۷۲	۱۰-۱۳. جمع‌بندی فصل

۳۷۳ پرسش‌ها ۱۰-۱۴

۳۷۷ فهرست اختصارات

۳۸۵ کتابنامه

www.ketab.ir

پیشگفتار

سیگنال گفتار به عنوان یکی از متداول ترین ابزار ارتباطی بین انسان ها و نیز بین انسان و ماشین مطرح است. بر این اساس، پردازش سیگنال گفتار به یکی از زمینه های فعال پژوهشی و کاربردی در نیم قرن اخیر تبدیل شده است. از زمینه های کاربردی سامانه های پردازش سیگنال گفتار می توان به کد کردن گفتار (با هدف فشرده سازی و رسیدن به یک بازنمایی با نرخ بیت پایین تر در سامانه های انتقال صوت)، بازشناسی خودکار گفتار، تصدیق/تعیین هویت گوینده، بازشناسی خودکار حالت احساسی گوینده، تشخیص خودکار نوع زبان، تبدیل متن به گفتار، بهسازی سیگنال گفتار و تبدیل صدا اشاره داشت.

در این کتاب به مرور مبانی پردازش رقمی سیگنال گفتار و نیز کاربردهای بازشناسی در قالب ۱۰ فصل پرداخته می شود. در راستای مرور مبانی این حوزه، ابتدا پس از معرفی ابتدایی موضوع و سامانه های کاربردی مختلف مربوط در فصل اول، موضوع تولید و درک سیگنال گفتار در فصل دوم مطرح می شود. سپس تحلیل های کوتاه مدت سیگنال گفتار در فصل سوم معرفی می شوند. تحلیل همومرفیک^۱ سیگنال گفتار در قالب معرفی ضرایب کپستروم^۲ کوتاه مدت نیز در فصل چهارم انجام می شود. فصل پنجم نیز به معرفی یکی از معروف ترین روش های تحلیل سیگنال گفتار (یعنی تحلیل پیش بینی خطی) می پردازد. این مبانی در اغلب کتاب های لاتین پردازش سیگنال گفتار آورده شده است. در این کتاب، از مبانی ارائه شده در کتاب های تألیفی آقایان ال. ر. رابینر^۳ و ر. دبلیو. شیفر^۴ در این مورد (مراجع [۳، ۱۲۸]) برداشت شده و مباحث فصول دوم تا پنجم شکل گرفته است. البته زمینه های به کارگیری این مبانی در سامانه های نوین پردازش گفتار به مطالب این فصول اضافه شده است. به عنوان نمونه، در فصل چهارم موضوع به کارگیری ضرایب کپستروم و نسخه های اصلاحی آن در کاربردهای بازشناسی گفتار، بازشناسی گوینده، بازشناسی حالت احساسی، بازشناسی نوع زبان، بازشناسی سن و جنس گوینده، سنتز و بهسازی گفتار، و آسیب شناسی گویش با توجه به پژوهش های دهه اخیر ارائه شده است.

در راستای معرفی سامانه های کاربردی پردازش گفتار، در فصل ششم انواع سامانه های کد کردن گفتار در قالب گستره وسیعی از روش ها مشتمل بر مدل های مبتنی بر شکل موج، مدل های پارامتری و مدل های کدگذاری ادراکی معرفی می شوند. کاربردهای بازشناسی نیز در فصول هفتم تا دهم کتاب معرفی می شوند. در تدوین این فصول نیز از مقاله های مروری هر یک از کاربردهای بازشناسی مذکور بهره گرفته شده

است. بدین ترتیب که سامانه‌های بازشناسی خودکار گفتار با تمرکز بر تحلیل ویژگی‌های گفتار با هدف بازنمایی مقاوم، روش‌های مختلف برای فرایند تصمیم‌گیری، بازشناسی گفتار طبیعی و دارای حالت، و روش‌های مقاوم‌سازی در برابر نویز و تغییرات محیط در فصل هفتم معرفی خواهند شد. در ضمن، در این فصل تلاش‌های پژوهشی و توسعه‌ای انجام‌شده در کشور برای بازشناسی خودکار گفتار فارسی به تفصیل ذکر خواهد شد.

در ادامه، سامانه‌های بازشناسی گوینده و نیز جنسیت و سن گویندگان در فصل هشتم و با تمرکز بر روش‌های مقاوم برای مقابله با شرایط نویز و عدم تطبیق نشست‌های آموزش و آزمون، دسته‌بندی مبتنی بر ابربردار، روش‌های جبران‌سازی، ذکر سامانه‌های نمونه و نیز مرور پژوهش‌ها در کشور مطرح می‌شوند. سامانه‌های بازشناسی حالت احساسی گوینده، موضوع فصل نهم این کتاب است که در آن ضمن بررسی موضوع طراحی دادگان گفتار دارای حالت احساسی، روش‌های مختلف استخراج ویژگی و دسته‌بندی ذکر و درنهایت، پژوهش‌های نوین انجام‌شده در این زمینه معرفی می‌شوند. سامانه‌های بازشناسی خودکار نوع زبان نیز در فصل دهم با تمرکز بر اطلاعات قابل به‌کارگیری در این سامانه‌ها (مشمول بر اطلاعات آکوستیکی، آوایی، نواپی، ریخت‌شناسی لغوی و دستورزبان)، روش‌های مختلف دسته‌بندی، دادگان معروف در این کاربرد و درنهایت مقایسه عملکرد برخی از ساختارهای نوین در این زمینه معرفی خواهند شد.

با توجه به اقبال زیاد در سه دهه اخیر به این حوزه تخصصی از مهندسی برق و کامپیوتر در کشور از یک‌سو، و کمبود منابع فارسی در این حوزه، سعی شده که در کتاب حاضر برخی از زمینه‌های کاری (به‌ویژه کاربردهای بازشناسی) با عمق مناسب مورد بررسی قرار گرفته تا جوابگوی نیاز در سطح یک مرجع درسی و نیز راه‌اندازی علاقمندان در انجام پژوهش و توسعه در این زمینه‌ها باشد. در تدوین این کتاب به موارد زیر توجه شده که به غنای کار افزوده شود:

الف- بیان مبانی و کاربردها به ترتیبی که کتاب به‌عنوان یک منبع برای درس **پردازش گفتار** در حوزه مهندسی برق و دروس مشابه در حوزه‌های مهندسی کامپیوتر، مهندسی پزشکی و مهندسی مکاترونیک قابل استفاده باشد. بدین منظور از قریب به ۱۳۰ شکل و ۳۰ جدول تفصیلی و نیز بالغ بر ۵۶۰ پرسش در ساختار کتاب استفاده شده است. در انتهای فصول مربوط نیز مسائلی برای حل، که به فهم موضوع کمک می‌نمایند، آورده شده است.

ب- ذکر روش‌های نوین در کنار مبانی مطرح در هر موضوع به به‌کارگیری بالغ بر ۷۰۰ مرجع که قریب به ۱۱۰ مرجع مربوط به سال‌های ۲۰۱۱ میلادی به بعد می‌باشد.

پ- ذکر فعالیت‌های پژوهشی و توسعه‌ای انجام‌شده توسط اساتید و پژوهشگران ارزشمند کشور در سه دهه اخیر در زمینه‌های مربوط با ذکر بالغ بر ۳۵ مرجع نمونه از سامانه‌های ارائه‌شده.

ت- ذکر قریب به ۳۰ مرجع از کارهای مربوط انجام‌شده توسط نویسندگان و همکاران در حوزه‌های مربوط در دو دهه اخیر، که اغلب مبتنی بر پردازش هوشمند سیگنال گفتار به کمک الگوریتم‌های هوش محاسباتی می‌باشد.

در نهایت، از پشتیبانی و صبوری همسر و فرزندانم در آماده‌سازی این کتاب و نیز پیشنهادهای سازندهٔ داوران محترم و ویراستار کتاب کمال سپاس را دارم. از حمایت‌های بی‌دریغ دفتر انتشارات علمی حوزهٔ معاونت پژوهشی و برنامه‌ریزی دانشگاه آزاد اسلامی - واحد تهران جنوب نیز نهایت تشکر را دارم. امید است که با توجه به کمبود منبع فارسی در این حوزه از یک‌سو و اهمیت سامانه‌های کاربردی این حوزه از سوی دیگر، کتاب حاضر مورد استفادهٔ علاقمندان قرار گیرد.

منصور شیخان

دانشیار گروه مهندسی مخابرات

دانشگاه آزاد اسلامی - واحد تهران جنوب

۱۳۹۳

www.ketab.ir